WATER RESOURCES RESEARCH, VOL. ???, XXXX, DOI:10.1029/,

An Objective Methodology for Merging Satellite- and Model-Based Soil Moisture Products

M. T. Yilmaz,[1] W. T. Crow,[1] M. C. Anderson,[1] and C. Hain[2]

[1]USDA, Hydrology and Remote Sensing Lab, Beltsville, MD, USA.

[2]University of Maryland, Earth System Science Interdisciplinary Center, College Park, MD, USA.

Abstract. An objective methodology, that does not require any user-defined parameter assumptions, is introduced to obtain an improved soil moisture product along with associated uncertainty estimates. This new product is obtained by merging model-, thermal infrared remote sensing-, and microwave remote sensing-based soil moisture estimates in a least squares framework, where uncertainty estimates for each product are obtained using triple collocation. The merged product is validated against in-situ based soil moisture data and showed higher correlations with observations than individual input products. The resulting combined soil moisture estimate is an improvement over currently available soil moisture products due to its reduced uncertainty and can be used as a stand alone soil moisture product with available uncertainty estimates.

1. Introduction

Consistent estimates of soil moisture can be obtained in various ways; for example through remote sensing or through modeling of the land-surface water budget. However, these estimates are not perfect and each method has characteristic uncertainties. Therefore, it is frequently desirable to merge independent realizations to obtain a more accurate unified estimate. Theoretically, the more independent data that are merged, the larger the reduction in the noise of the merged product. However, it is important to weight the products based on their relative accuracies in order to minimize errors.

Data assimilation using Kalman Filter-based methodologies is one of the most commonly-used approaches for merging different products while taking into account the relative uncertainties. However, in land data assimilation studies, these methodologies often rely on ad-hoc statistical descriptions of errors in assimilated observations, model parameters, or model forcings. As a result, the relative weighting applied to modeled and observed soil moisture information by a land data assimilation is arguably subjective and does not necessarily reflect an optimized integration of independent data sources [Crow and Van Loon, 2006; Reichle et al., 2008].

Kalman Filter theory can be shown to be a recursive solution of the least squares problem [Sorenson, 1970] for an appropriate time frame. The solution of Kalman [1960] enables propagation of the best estimate and its errors in time, whereas in ordinary least squares the solution is assumed constant in time. The ultimate goal for both of these solutions can be shown to obtain an estimate that has minimized error variance. However, both solutions require prior knowledge of the uncertainty estimates of the products to obtain an optimal analysis.

Triple collocation is a method that objectively obtains error estimates for three or more independent products. This method was originally introduced in oceanic studies by Stoffelen [1998] and Caires and Sterl [2003] to estimate near-surface wind speed errors, and later applied in many hydrological applications. Scipal et al. [2008] estimated the errors in passive microwave-, active microwave-, and model-based soil moisture products. Miralles et al. [2010] estimated errors in passive microwave-, station-, and model-based soil moisture products and validated the error estimates using watershed scale station-based data. Dorigo et al. [2010] evaluated the uncertainties of global passive microwave-, active microwave-, and model-based soil moisture products. Hain et al. [2011] estimated errors in passive microwave-, thermal infrared-, and model-based soil moisture realizations. Parinussa et al. [2011b] estimated errors in passive microwave-, active microwave-, and antecedent precipitation index-based soil moisture products, compared the triple collocation-based errors with data assimilation based error estimates [Crow, 2007], and found very high correlation between the error estimates of these two techniques.

Triple collocation was advocated by Crow and van den Berg [2010] as a means to estimate observation error covariance parameters required by land data assimilation systems. However, Crow and van den Berg [2010] were still forced to make a number of subjective guesses regarding the statistical attributes of modeling error in their system. In this study, we propose an objective methodology that does not require any user-defined error parameters as input. In this approach, different soil moisture products are merged in a least squares framework that relies on the error estimates of the products obtained from triple collocation. Specifically, we merge thermal remote sensing based soil moisture proxy retrievals from the Atmosphere Land Exchange Inversion [ALEXI; Anderson et al., 2007a] energy balance model, the Noah [Ek et al., 2003] land surface model (LSM) soil moisture simulations, and Land Parameter Retrieval Model [LPRM; Owe et al., 2008] soil moisture estimates based on microwave remote sensing observations. The least squares framework is also able to provide estimates of uncertainty in the merged product. The methodology proposed here can potentially add value to the soil moisture products derived from the current and future soil moisture satellite missions (i.e, SMOS: Soil Moisture and Ocean Salinity; SMAP, Soil Moisture Active Passive) by optimally merging them with independent soil moisture estimates acquired from infrared observations and land surface models.

The general least squares solution is briefly reviewed in the next section. Section 3 reviews the triple collocation equations, section 4 introduces the input data, section 5 presents the results, and section 6 summarizes our conclusions.

2. Least Squares Merging

Least squares is an estimation theory that has been used in numerous studies since its initial applications by Gauss [1963] and Legendre [1806]. The theory has gained its current form by Kalman [1960] [Sorenson, 1970] and can be used to describe the basis of most modern data assimilation techniques [Talagrand, 1997]. We use least squares to optimally merge multiple independent products with known uncertainty estimates. The least squares solution has been derived in many studies; here we briefly review it to provide background for our proposed merging algorithm.

Assuming we have three independent realizations (S_x, S_y, and S_z) of a variable along with their respective zero-mean errors (ϵ_x, ϵ_y, and ϵ_z) and error variances (σ_x^2, σ_y^2, and σ_z^2). These realizations can be represented by

$$S_x = \alpha S_t + \epsilon_x \tag{1}$$

$$S_y = \alpha S_t + \epsilon_y \tag{2}$$

$$S_z = \alpha S_t + \epsilon_z \tag{3}$$

where S_t is the true value of the variable and α is a measure of the relation between these realizations and the assumed truth. Although in some cases $\alpha = 1$, this is not a requirement; the least squares solution can be obtained as long as all realizations relate to the truth with the same coefficient. The desired merged estimate, S_m, is obtained as

$$S_m = w_x S_x + w_y S_y + w_z S_z \tag{4}$$

where w_x, w_y, and w_z are the relative weights of S_x, S_y, and S_z respectively. To have an unbiased merged estimate ($E[S_m - \alpha S_t] = 0$), it is required that $w_x + w_y + w_z = 1$. Given these constraints, the ultimate goal is to derive these weights as functions of the error variance of the three realizations and to find the error variance estimate of the merged product. The error estimate of the merged product is obtained as $\epsilon_m = S_m - \alpha S_t$ and the solution we seek minimizes a selected cost function (J) in a mean squares sense. Here, we select this cost function to be the error variance of the merged estimate:

$$J = \sigma_m^2 = w_x \sigma_x^2 + w_y \sigma_y^2 + w_z \sigma_z^2 \tag{5}$$

$$J = \sigma_m^2 = w_x \sigma_x^2 + (1 - w_x - w_z)\sigma_y^2 + w_z \sigma_z^2. \tag{6}$$

Setting $\partial J/\partial w_z = 0$ and $\partial J/\partial w_x = 0$ in eq. 6 and solving for w_x, w_y, and w_z, we obtain

$$w_x = \frac{\sigma_y^2\sigma_z^2}{\sigma_x^2\sigma_y^2 + \sigma_x^2\sigma_z^2 + \sigma_y^2\sigma_z^2} \tag{7}$$

$$w_y = \frac{\sigma_x^2\sigma_z^2}{\sigma_x^2\sigma_y^2 + \sigma_x^2\sigma_z^2 + \sigma_y^2\sigma_z^2} \tag{8}$$

$$w_z = \frac{\sigma_x^2\sigma_y^2}{\sigma_x^2\sigma_y^2 + \sigma_x^2\sigma_z^2 + \sigma_y^2\sigma_z^2}. \tag{9}$$

The solution is intuitive since the weights are proportional to the uncertainty of the other two estimates. If two realizations are available instead of three, then the least squares solution can be applied similarly with a cost function selection of

$$J = \sigma_m^2 = w_x\sigma_x^2 + (1 - w_x)\sigma_y^2 \tag{10}$$

where the weights are obtained as

$$w_x = \frac{\sigma_y^2}{\sigma_x^2 + \sigma_y^2} \tag{11}$$

$$w_y = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_y^2}. \tag{12}$$

The merged product at any given time can therefore be based on anywhere between one and three realization(s). Hence, the uncertainty of the merged product at any given location may not be constant in time. Accordingly, for each available merged product, its uncertainty is also given as a separate product. The alternative is to use only the mutually available data to preserve the uncertainty estimate of the merged product in time. However, in this latter scenario, temporal and spatial gaps of the merged product would be larger and the merged product would have higher uncertainty.

3. Error Estimation Using Triple Collocation

For a given set of realizations, optimal merging based on least squares technique described here requires an estimate of the relative uncertainties of input products. In this study, the error variances of these estimates are obtained using triple collocation. Triple collocation is an attractive methodology that estimates the relative errors of different products regardless of their observation platform. Triple collocation solutions were first introduced in oceanic applications by Stoffelen [1998] and Caires and Sterl [2003], and later applied in many hydrological studies [Parinussa et al., 2011b; Loew and Schlenz, 2011]. From now on, we use the abbreviations ST1998 and CS2003 to refer to the triple collocation solutions introduced by Stoffelen [1998] and by Caires and Sterl [2003] respectively. The ST1998 is flexible enough to accommodate representation errors (i.e. point vs grid data), whereas this component is neglected in CS2003. On the other hand, CS2003 accommodates correlated errors between realizations, whereas error cross-correlations are required to be zero in the solution of ST1998. Moreover, ST1998 explicitly requires a rescaling step to enforce datasets to have the same relationship with the truth. CS2003 does not require this rescaling, and as a result the error variance estimates obtained before and after a potential rescaling (if applied) differ. If this rescaling step is applied, both CS2003 and ST1998 yield identical error variance estimates under same assumptions.

Given that the ultimate goal of this study is to merge different estimates, it is necessary to rescale them to obtain a set of realizations that has consistent relationship with the assumed truth, similar to eq. 1–3. Hence, we adopt ST1998 in this study:

$$S_1 = \alpha_1 S_t + e_1 \tag{13}$$

$$S_2 = \alpha_2 S_t + e_2 \tag{14}$$

$$S_3 = \alpha_3 S_t + e_3 \tag{15}$$

where S_t is the true soil moisture anomaly with variance σ_t^2; S_1, S_2, and S_3 are three soil moisture anomalies related to truth with α_1, α_2, and α_3 coefficients, with errors e_1, e_2, and e_3, and with error variances σ_1^2, σ_2^2, and σ_3^2 respectively. Here σ_t^2 does not imply the truth has errors, but rather it is the true soil moisture variance in time. We rescale these realizations using:

$$S_1^* = \alpha S_t + e_1^* \tag{16}$$

$$S_2^* = \alpha S_t + e_2^* \tag{17}$$

$$S_3^* = \alpha S_t + e_3^* \tag{18}$$

where S_1^*, S_2^*, and S_3^* are the rescaled realizations and e_1^*, e_2^*, and e_3^* are the relative errors of the realizations with variances ${\sigma^*}_1^2$, ${\sigma^*}_2^2$, and ${\sigma^*}_3^2$. Rescaled values are related to the initial estimates as $S_1^* = S_1 c_1$, $S_2^* = S_2 c_2$, and $S_3^* = S_3 c_3$, where c_1, c_2, and c_3 are the rescaling factors. By arbitrarily selecting any of the datasets as a reference (in this study assuming $\alpha = \alpha_1$) and assuming the representativeness errors that are described by Stoffelen [1998] are zero, the rescaling factors can be found as,

$$c_1 = 1 \tag{19}$$

$$c_2 = \frac{\overline{S_1^* S_3^*}}{\overline{S_2^* S_3^*}} \tag{20}$$

$$c_3 = \frac{\overline{S_1^* S_2^*}}{\overline{S_3^* S_2^*}}. \tag{21}$$

Error variance estimates (σ_1^2, σ_2^2, or σ_3^2) for the original non-scaled datasets (S_1, S_2, and S_3) using CS2003 can be converted into the error variance (${\sigma^*}_1^2$, ${\sigma^*}_2^2$, or ${\sigma^*}_3^2$) of the scaled estimates (S_1^*, S_2^*, and S_3^*) using the same rescaling factors given in (19–21). However, it is emphasized that applying ST1998 without the rescaling step does not necessarily give the error variances of the non-scaled datasets as opposed to applying CS2003. Additionally, the climatologies are removed with the standardization process so that datasets have zero mean (consistent with ST1998 and CS2003) and unity standard deviation. Consequently, the TC analysis is performed solely on soil moisture anomalies and is not impacted by the likely presence of bias in one or more of the datasets. Here the accuracy of the rescaling to match the relations of the datasets with the truth is tied to the linear relation between the products in the form given in (1-3). When compared to more nonlinear systems, highly linear systems are expected to have smaller sampling errors and require fewer observations to obtain same level of accuracy. Also, note that this rescaling step can be performed independently for each area or time period of interest, hence it may vary spatially or temporally.

In the triple collocation system of equations presented above, there are current seven unknowns (α_1, α_2, α_3, σ_t^2,σ_1^2, σ_2^2, and σ_3^2) constrained by three equations (16–18). By selecting a reference dataset (i.e. assuming $\alpha = \alpha_1$) and rescaling other datasets to this reference, our goal becomes seeking a solution for four unknowns ($\alpha^2\sigma_t^2$, ${\sigma^*}_1^2$, ${\sigma^*}_2^2$, and ${\sigma^*}_3^2$), rather than seven. This system, with four unknowns and three equations, is still under-determined. We are able to solve for these four unknowns only after assuming all error related cross-covariances are zero.

However, in the absence of any other independent information, we cannot decompose the $\alpha^2\sigma_t^2$ estimate into estimates of α^2 and σ_t^2; meaning we can never know the true σ_t^2. Different reference dataset selections result in different $\alpha^2\sigma_t^2$ as well as different ${\sigma^*}_1^2$, ${\sigma^*}_2^2$, and ${\sigma^*}_3^2$. Therefore the triple collocation equations described above provide only the relative accuracy of these realizations (how the noisiness of one product compares against other product) whereas the absolute values of the error variances themselves are dependent on the reference dataset selection. While triple collocation is not ideal for capturing absolute errors, its representation of relative errors between input products is independent of the arbitrary choice of a single dataset as a scaling reference. Fortunately, this type of relative information - and not absolute errors - is all that is required in order to determine optimal least-squares averaging.

Assuming the errors of these products are independent from each other and from the truth, and assuming a mutual linear relationship between these estimates and the true soil moisture, the final error variances of the rescaled realizations (that are used in the above described least squares solution) are found as:

$$\sigma^{*2}_1 = \overline{(S_1^* - S_2^*)(S_1^* - S_3^*)} \tag{22}$$

$$\sigma^{*2}_2 = \overline{(S_2^* - S_1^*)(S_2^* - S_3^*)} \tag{23}$$

$$\sigma^{*2}_3 = \overline{(S_3^* - S_1^*)(S_3^* - S_2^*)}. \tag{24}$$

Note that the triple collocation error variances, which are assumed constant in time, are estimated using the entire time series only when at least 100 separate retrievals/estimates are mutually available for each of the 3 input soil moisture products. If this threshold is not met, then all error variance estimates are assumed equal (i.e., triple collocation is not calculated). Once these error variance estimates are obtained, weights are calculated at each time step independently using these error variances in a least squares framework. While the obtained error variance estimates are constant in time, the weights are not. When all three realizations are available, the least squares solution for three datasets (eq. 7–9) is used; when two out of three realizations are available then the least squares solution for two datasets (eq. 11–12) is used.

Accordingly, the error variance of the merged product at each time step is calculated using eq. 6 or eq. 10, depending on the number of available realizations at any given time. When only one realization is available, this single product is used as the final merged product and its error variance is used as the error variance of the merged product. If all realizations would have had the identical temporal coverage (all available or all missing simultaneously), then the weights would have been constant in time. They change in time only due to the availability of the products at any given time step. Then the datasets are merged using these calculated weights for each time step separately. If there are not enough mutually available products, meaning a triple collocation based estimate is not available, then products are merged using equal weights.

4. Data

4.1. Input Datasets

The study area is selected as the continental United States, between 125°-67°W and 25°-50°N. Daily datasets are obtained for each year from 2002 to 2010 for the months of April through October. Large-scale soil moisture information is currently available from three independent sources: retrievals derived from thermal-infrared remote sensing, retrievals derived from microwave remote sensing, and estimates derived from water balance models forced with micro-meteorological observations. Here, all three sources of soil moisture data are used as input into the triple collocation analysis. In particular, this study utilizes an ALEXI energy balance model soil moisture proxy obtained from thermal infrared remotely sensed images, LPRM soil moisture estimates that are obtained from passive microwave remote sensing images, and Noah land surface model soil moisture simulations. The methodology is applied at a grid space of 0.25°; datasets at higher native resolution have been aggregated to this common grid. All datasets are averaged to weekly composites from their native temporal resolution.

ALEXI is a two-source (soil and vegetation) model that solves for the latent heat and the sensible heat components of the surface energy balance by taking advantage of measurements of morning land-surface temperature rise obtained by geostationary satellites reducing sensitivity to absolute biases in retrieved temperature [Mecikalski et al., 1999; Anderson et al., 2007a]. Using the obtained fluxes, a strong relationship was found between the ratio of actual to potential evapotranspiration fluxes (also named as fraction of potential evapotranspiration; f_{PET}) and the fraction of available water (f_{aw}) in the soil column [Anderson et al., 2007a, b, 2011]. Following this study, Hain et al. [2009] proposed unique relationships between f_{PET} and f_{aw}, evaluated this relation using soil moisture observations from the Oklahoma Mesonet Network, and showed ALEXI has valuable information about f_{aw} which serves as a proxy for the root-zone soil moisture in the vegetated areas. Here we utilize ALEXI-based f_{PET} retrievals following the approach described by Hain et al. [2011]. Note that ALEXI f_{PET} represents a surface–root-zone merged soil moisture estimate; yielding a proxy estimate of water availability for evapotranspiration (i.e. water in the surface layer for bare soil evaporation, and water in the root-zone for canopy transpiration). ALEXI f_{PET} values have been aggregated from 10km to 0.25° resolution. Given its reliance on the thermal remote sensing based observations, current ALEXI retrievals are limited to clear-sky conditions, which is a major limitation to data availability particularly over the Northern US. To fill the entire grid, it is necessary to average daily f_{PET} fields over time to create time composites. More detailed information about ALEXI based soil moisture proxy can be found in above mentioned studies.

Noah (version 2.7) LSM data were obtained from the global simulations generated using Global Land Data Assimilation System [GLDAS; Rodell et al., 2004] forcing data. The Noah model calculates a coupled surface water and energy balance and thus calculates multi-layer soil moisture as the storage component of a soil water balance. More details about these Noah simulations can be found at http://disc.sci.gsfc.nasa.gov/hydrology/documentation. These hourly simulations were performed at 0.25° spatial resolution, hence spatial aggregation was not needed. Since the ALEXI soil moisture proxy has mixed vertical support over sparsely and densely vegetated surfaces, a Noah soil moisture estimate is computed that mimics this vertical support. The second-layer (10-40cm depth) and the third-layer (40-100cm depth) soil moisture simulations are averaged into a root-zone soil moisture estimate ($Noah_{root}$) by weighting each layer volumetric soil moisture proportional to respective soil layer depths. This root-zone product and the surface (0-10cm) soil moisture simulations ($Noah_{srfc}$) are later combined into an adjusted soil moisture estimate ($Noah_{adj}$) following the study of Hain et al. [2011]:

$$Noah_{adj} = (1 - f_{vc})Noah_{srfc} + f_{vc}Noah_{root} \tag{25}$$

where f_{vc} is the fractional vegetation cover based on remote sensing based observations of leaf area index acquired by the Moderate Resolution Imaging Spectroradiometer (MODIS). As a result of eq. 25, $Noah_{adj}$ estimates are essentially surface soil moisture estimates over areas with no vegetation cover, and are root-zone soil moisture estimates over areas with dense vegetation cover.

Advanced Microwave Scanning Radiometer EOS (AMSR-E) microwave remote sensing based brightness temperature observations have been used in numerous passive microwave-based algorithms [Jackson, 1993; Owe et al., 2001; Njoku and Chan, 2006; Lu et al., 2009], and the resulting soil moisture products have been extensively validated under a wide range of ground conditions and climate regimes [Draper et al., 2009; Mladenova et al., 2011; Parinussa et al., 2011a]. Among these products, LPRM soil moisture estimates have been used in this study [Owe et al., 2008], obtained from Vrije University Amsterdam (VUA). LPRM soil moisture estimates are obtained using one layer radiative transfer-based land parameter retrieval model. This retrieval model uses soil related information as ancillary data, and solves simultaneously for soil moisture, vegetation optical depth, and soil skin temperature. The model uses the relationship between Microwave Polarization Difference Index, vegetation optical depth, and soil dielectric constant and solves for the skin temperature using a regression-based model based on Ka-band vertical polarization AMSR-E brightness temperature data [Holmes et al., 2009]. Soil moisture retrievals are based on C-band descending AMSR-E brightness temperature observations. However, X-band observations are also used in areas of the world where C-band observations are affected by radio frequency interference. The LPRM soil moisture estimates refer to the top 3cm of the soil profile. AMSR-E-based brightness temperature (Tb) observations are obtained at native spatial resolutions of 56km and 38km for C- and X-band, respectively. The operational LPRM product has been re-gridded to 0.25° spatial resolution are re-gridded values by taking advantage of the multiple samples within the same footprint.

Here the three parent datasets are obtained from different algorithms driven by different input data, supporting the assumption of the independence of the errors for the triple collocation methodology. On the other hand, these products may have different skills in predicting the truth that we define. However, here it is stressed that as long as highly linear relationships exist between the products, the dataset selection does not present any problem in a triple collocation based framework regardless of the differences in the dataset retrieval algorithms. This issue will be revisited in the results section to provide more elaborate discussions.

In terms of timing, ALEXI provides a direct estimate of the soil moisture conditions at shortly before the local noon on days with clear morning conditions. LPRM soil moisture retrievals are obtained using microwave remote sensing based observations collected at 1.30am UTC. On the other hand, Noah SM estimates are temporally continuous, and output at hourly time-steps. Accordingly, there could be minor inconsistencies between the temporal representativeness of these products. However, the impact of these inconsistencies should be minimized during the temporal averaging to obtain weekly composites. Given orbit patterns and typical frequency of mask retrievals, ALEXI and LPRM weekly composites are obtained by averaging around 2-4 daily retrievals whereas Noah weekly composites are obtained by averaging 24*7=168 hourly simulations. Hence, Noah has better "weekly" temporal support than do the other products. However, it should be noted that poor support is simply one component of the total random error detected by triple collocation and therefore poses no particular challenge for our proposed merging strategy.

4.2. Validation Datasets

The merged product has been evaluated in comparison with in situ soil moisture observations from the Oklahoma MESONET Network [Brock et al., 1995; Basara and Crawford, 2000] and the Soil Climate Analysis Network [SCAN, Schaefer et al., 2007] within the Contiguous United States (CONUS). In Oklahoma an integrated network of 135 meteorological stations has been installed during the past two decades. Among these stations, around 100 have calibrated soil moisture monitoring devices taking measurements at 5cm, 25cm, 60cm, and 75cm depths. Collected data undergo automated and manual quality controls conducted by University of Oklahoma during the conversion of 30min raw data into daily soil moisture averages [Illston et al., 2008]. There are over 150 SCAN stations spread throughout the CONUS taking soil moisture measurements at 5cm, 10cm, 20cm, 50cm, and 100cm depths [Schaefer et al., 2007].

In a manner analogous to eq. 25, a vegetation correction has been applied to the station measurements to ensure consistent soil moisture estimates between the merged products and the validation datasets. More specifically, the 1st layer (top 5cm) MESONET data have been taken as surface soil moisture and a weighted average of the 2nd to the 4th layers as a root zone; the MODIS-based vegetation cover fraction information at 0.25 degree grid is assumed to be a representative value for the station location, where the vegetation correction is carried out using eq. 25. Similarly a vegetation correction was also applied to SCAN soil moisture values; the 1st layer soil moisture values are used as surface values and average soil moisture values of the 2nd to the 5th layers, weighted by their depths, are used as root zone values. The merged soil moisture estimates were validated using these vegetation-cover adjusted soil moisture observations. Because the MESONET and the SCAN station data are adjusted for vegetation cover fraction, the number of available station data points depends on the availability of both the surface and the root-zone observations. Since the root-zone observations are not as readily available as the surface observations, there are approximately only 50 MESONET and SCAN stations available for verification.

The skill of the triple collocation based weights was also evaluated by comparing the performance of merged estimate against the performance of a naively-merged product performance, which simply assumes equal weights for each available product.

4.3. Data Standardization

Weekly composites are standardized, so that their time-mean (across years) is zero and time-variance is unity for a given pixel and week.

$$\mu_{w,lon,lat} = \sum_{y}^{nar} SM_{y,w,lon,lat}/nar \tag{26}$$

$$\sigma_{w,lon,lat} = (\sum_{y}^{nar} (SM_{y,w,lon,lat} - \mu_{w,lon,lat})^2/nar)^{1/2} \tag{27}$$

$$SMs_{y,w,lon,lat} = \frac{SM_{y,w,lon,lat} - \mu_{w,lon,lat}}{\sigma_{w,lon,lat}} \tag{28}$$

where y, w, lon, and lat denote year, week, longitude, and latitude respectively; SM denotes one of the three soil moisture products used in this study (ALEXI, Noah, and LPRM); SMs is the standardized soil moisture realization; and nar is the number of available realizations out of 9 years for the given week, longitude, and latitude. The standardized SMs values defined above were used in the triple collocation based error estimations. Here the merging process could have been performed by adjusting for only the mean component of the products; however, standardization facilitates a more meaningful product comparison between the parent products and the merged product (with similar soil moisture magnitudes).

4.4. Vertical Support

The output product produced by the merging methodology introduced above is a surface–root-zone merged soil moisture estimate representing a proxy estimate of water available for evapotranspiration. The vertical support in each parent product, however, is different. ALEXI and Noah soil moisture represents a mixture of surface and root-zone moisture content, while the LPRM data reflect only surface (zero to 3-cm) soil moisture information and therefore has a different vertical support than Noah and ALEXI soil moisture products over vegetated areas. The effect of this inconsistency in vertical support over vegetated areas is investigated further by applying additional triple collocation analyses to vegetation-adjusted LPRM values that are obtained using an exponential filter methodology parameterized by various characteristic length scales and by examining the Noah and Common Land Model (CLM; see below for the description of CLM) correlations between surface and vegetation-adjusted soil moisture values.

Additional triple collocation analyses were performed using vegetation-adjusted LPRM values obtained using eq. 25. This equation uses the native LPRM surface and LPRM-based root-zone products obtained using the exponential smoothing methodology described by Wagner et al. [1999] and Albergel et al. [2008] to estimate root-zone soil moisture retrievals from superficial observations:

$$LPRM_{root}(t) = \frac{\sum_i LPRM_{srfc}(t_i) e^{-(t-t_i)/\tau}}{\sum_i e^{-(t-t_i)/\tau}} \tag{29}$$

where $t_i \leq t$, $LPRM_{srfc}$ is the surface LPRM soil moisture estimate at time t_i, $LPRM_{root}$ is the root-zone soil moisture estimate, and τ is the characteristic time length. Specifically three vegetation-adjusted LPRM products were estimated using three separate root-zone LPRM values obtained via assigning τ values of 4, 7, and 14 days. Accordingly, we have performed four parallel triple collocation analyses that use the same ALEXI and Noah datasets but different LPRM-based soil moisture values (one LPRM-surface product and three vegetation-adjusted LPRM products).

In this study we also use CLM (version 2.0) simulations, solely for the investigation of surface–vegetation-adjusted soil moisture values coupling and not in the triple collocation merging methodology (section 5 below). Like Noah, CLM is a soil-vegetation-atmosphere transfer model that solves for the water and the energy balance at the surface [Dai et al., 2003], and is driven here by GLDAS forcing data [Rodell et al., 2004]. CLM simulations have 1° spatial resolution and utilize 10 soil layers with 2, 3, 4, 8, 12, 20, 34, 55, 92, and 113cm depths respectively. Vegetation-adjusted CLM soil moisture values were obtained (25) by using surface soil moisture estimates defined as the weighted average of the 1st to the 3rd layers (0-9cm) and using root-zone soil moisture estimates defined as the weighted average of the 4th to 7th layers (10-83cm).

4.5. Additional Considerations

For cross-comparisons of the linear relation between parent products, cross-correlations were calculated without setting any threshold for the availability of the products. The resulting correlation values were then masked if a significant correlation was not found. For the triple collocation we have set a minimum number (100) of mutually available datasets. If 100 mutually available soil moisture values were not found, then the triple collocation analysis was not performed. In such cases, 0.33 weights are assigned for all three products. However, for the data merging on each individual date, the actual weights depend on the availability of the datasets for that particular day. For example for a pixel that has equal weights, if all three datasets are available for any given day, only then equal weights are used; if only two of the products are available at any given day, then the applied weights would be recalculated to 0.50 and 0.50.

Triple collocation based error estimates are also dependent on the availability of the daily products, which influences the uncertainty of the sampled weekly composites. The more frequently a dataset is available, the less noisy its weekly composite become. On average ALEXI has 2.1 and LPRM has 3.1 available observations per week over the CONUS, whereas Noah weekly estimates are based on 168 separate hourly Noah soil moisture predictions generated each week (i.e., 24 estimates/day times 7 days). Although it is possible to combine both the ascending and the descending AMSR-E based LPRM soil moisture estimates to increase the number of mutually available observations, this has not been done in this study. Here, it should be noted that the merged weekly composite is derived from the weighted averaging of either one, two or three individual soil moisture products. Hence, the uncertainty of the final merged estimate at any week also depends on the availability of the products. Dates with more missing soil moisture values have higher uncertainty compared to dates with less missing values.

5. Results

5.1. Correlations and Weights

ALEXI, Noah, and LPRM based soil moisture anomaly estimates were used to calculate the error variances of each product in a triple collocation framework. As triple collocation based error estimates require a mutual linear relationship between products, we have evaluated the linearity between the three products by analyzing their cross-correlations (Fig. 1). Significant correlations between LPRM and ALEXI, and between LPRM and Noah over large parts of the eastern CONUS are not found, which is partly due to the non-availability of LPRM soil moisture estimates caused by the strong attenuation of the microwave signal over densely vegetated areas. On the other hand, there are strong cross-correlations over areas of the southern and the northern CONUS (i.e. from Texas to Montana), indicating a strong mutual linear relationship between various soil moisture products.

The triple collocation based errors were computed using eq. 22–24 and were used in the least squares framework to obtain weights using eq. 7–9. In general, the differences between triple collocation analyses that use different LPRM products (corresponding to various amounts of temporal smoothing via eq. 29) are minimal (Fig. 2), suggesting the nonlinearities due to vertical support differences do not have a major impact on estimated weights, even though the use of longer exponential filter correlation lengths favor ALEXI more than Noah and LPRM with respect to the difference between the top and the bottom rows in Fig. 2. The resulting weights shown in Fig. 2 are intuitively consistent with the cross-correlations of the products (Fig. 1); the product that has the highest cross-correlation with its pairs also has the largest estimated weights. For example, the correlations between Noah and ALEXI and between Noah and LPRM are higher than the correlation between ALEXI and LPRM over the south-eastern CONUS; therefore, Noah weighting is relatively higher than both ALEXI and LPRM over this area (top row in Fig. 2). Similarly, the correlations between LPRM and ALEXI and between LPRM and Noah are higher than the correlation between ALEXI and Noah over the northern CONUS; therefore, the optimal weighting applied to LPRM retrievals is higher than ALEXI and Noah over this area. In general, ALEXI performs better over the southern CONUS than the northern, which can be attributed to the lower temporal coverage of ALEXI over the northern CONUS due to clouds [Hain et al., 2011].

This study focuses on the warm season to avoid issues related to snow cover and frozen soils, although it is possible to perform the analysis using both the warm and the cold season data. In general we may expect remote sensing based soil moisture estimates retrieved during winter to have higher sampling errors due to larger data gaps (both temporally and spatially) partly caused by snow and ice conditions than estimates retrieved during summer. Hence, a single set of weights for the entire year may not reflect the error characteristics as well as seasonally derived weights. The estimation of seasonal weights, however, would require longer time series and may be feasible with ongoing efforts to extend the length of the remote sensing-based databases.

5.2. Merged Estimate and Station Data

All subsequent merging results are based on the case of no LPRM smoothing (i.e., the top row in Fig. 2). For the merging methodology, the weights in Fig. 2 are used only when all three the datasets are available; for missing days, weights were calculated using the error estimates of the available days. Parent products (ALEXI, Noah, and LPRM), the merged estimate (merged realization using least squares) and the uncertainty of the merged estimate for the 19th week (7-13th of May) of 2007 are shown in Fig. 3. In this particular week, the standard deviation of the error estimate is around 0.40 (unitless as all products are standardized), and the soil moisture anomalies range between -2.6 to +2.7 standard deviations around the climatology of the given local pixel.

Time series of the parent products and the merged estimate are shown together with data from two individual MESONET and SCAN stations in Fig. 4. The weights of the parent products are similar at these station points; hence, the merged estimates fall between three parent products without closely following any one in particular. Average station data correlations with the parent products and the merged estimate are summarized in Table 1; the significance of these correlations, the correlation comparisons of parent products, and the merged estimate are given in Table 2. On average, parent products are better correlated with the MESONET data than the SCAN data (upper sections of Table 1). The number of stations that have significant correlations with the parent products and the merged estimate are higher for the MESONET data than the SCAN data (upper sections of Table 2). The merged estimates are better correlated with the station data than the individual parent products (middle sections of Table 1), particularly better than both ALEXI and LPRM (middle sections of Table 2), implying the merged product is more accurate than its parents products individually. Although on average the merged estimate has better correlation with the MESONET (but not SCAN) than the best correlation of the parent products, the improvement is not significant for the majority of the stations (lower sections of Table 2).

5.3. Implications of Naive Merging

Although application of the merging scheme leads to an integrated product that was generally better than any of its three parent products in isolation, the triple-collocation based merge estimate did not generally lead to an integrated product that was demonstrably superior to naive aggregation (i.e., aggregation with equal weighting) (Table 2). Potential reasons for the lack of significant improvement against the parent and the naively merged products include: 1) station data are point data and may have high representativeness errors [Ryu and Famiglietti, 2005; Miralles et al., 2010; Cosh et al., 2006], and/or 2) triple collocation based errors may not be optimum due to inadequate mutually available data (limited temporal extent of parent products), and/or 3) the weights are optimum, but the parent products may have similar skills and therefore merging them in a naive way produces estimates that are only marginally different from the optimally merged estimates obtained via triple collocation.

In particular the station observations are point data, thus very susceptible to representativeness errors and the weights obtained through triple collocation are very sensitive to the length of the mutually available data. It is our experience that the number of mutually available triplets in this study may not be sufficient for highly accurate triple collocation estimates on weekly or monthly time-scales. However, as longer time-series become available through remote sensing techniques and modeling, and as improved better station data (with less representativeness errors via better selection of station and/or sensor locations) are collected, it is expected that the merged estimates will result in higher improvements over the parent products.

The difference between the optimal solution and the naive method was also evaluated by investigating the sensitivity of the optimal solution to data availability and averaging. Specifically, the triple collocation based weights and the cross-correlations for various averaging windows-lengths were calculated (Table 3) to evaluate the sensitivity of derived optimal weights to aggregation period and retrieval availability. To do this, the daily data were averaged into either weekly or monthly composites, and using all the available daily data for averaging (i.e. the "all available scenario") or using only the days when all three products are available (i.e. the "mutually available scenario"). Applying the mutually available scenario guarantees that equal numbers of daily products are used in weekly or monthly composites analyzed via triple collocation. In general, the differences in weights were higher than the differences between cross-correlations for weekly all available scenario and the weight differences were much less for the weekly mutual scenario and for both monthly scenarios (Table 3). This implies that the weighting favors products with higher temporal availability (=model) for weekly scenarios, but the effect of this retrieval frequency is reduced when datasets are averaged for longer time periods. This reduced difference in weights and correlation can explain the similarity between the performance of merged products based on triple collocation and naive weighting. The skills of the parent products are very similar; therefore, the naive averaging approach simply follows the optimal solution obtained via triple collocation.

5.4. Vertical Support

As discussed above, the final merged soil moisture estimate is a mixed product that reflects the soil moisture layer that is actively interacting with the atmosphere via evapotranspiration. Hence, using the surface-only microwave remote sensing product over sparsely vegetated areas is consistent with the properties of the mixed product. However, over densely vegetated areas this mixed vertical support is inconsistent with microwave-based soil moisture retrievals, which are strictly limited to the near-surface layer (surface to 3cm). Consequently, over densely vegetated areas there is a potential inconsistency in the vertical support of LPRM soil moisture retrievals relative to ALEXI and Noah products (see above). A series of analyses has been performed to test the effect of using surface-only microwave remote sensing product on our triple collocation results over vegetated areas.

Since the parameter of interest is the vegetation-adjusted soil moisture value (rather than root-zone soil moisture), we have narrowed our focus to this parameter. High correlations at weekly time scales over densely vegetated areas imply a strong linear relation between the surface and the vegetation-adjusted soil moisture simulations; similar to the triple collocation equations (eq. 16-18) where we assume a linear relation between each dataset and the truth. Therefore the applicability of these equations to soil moisture products obtained at different vertical depths is determined by the linearity of the relationship between surface and vegetation-adjusted soil moisture. The depth variations pose a problem to our approach only if they manifest themselves in a nonlinear or a hysteric relationship between products. Conversely, if the relationship is linear, it simply folds into the linear rescaling step which underlies the application of triple collocation. Therefore the impact of vertical consistency (between LPRM and Noah/ALEXI-based soil moisture products) will hinge on the degree to which soil moisture estimates at various depths can be linearly related.

Correlations were computed between the surface and vegetation-adjusted soil moisture values from both Noah and CLM LSMs (Fig. 5) and both MESONET and SCAN station data (Table 4). Very high correlations (i.e., linear relationships) were found between the surface and the vegetation-adjusted station-based soil moisture data from station-based analysis (in Table 4, 0.91 for both MESONET and SCAN data) and from model simulations (in Table 4, 0.96 and 0.92 correlations for Noah and CLM respectively). Depending on these very strong linear relations between the surface and the vegetation-adjusted soil moisture values, we can tell with high confidence that -at weekly time scales- vertical inconsistencies in support can be effectively resolved via linear rescaling.

Another way to test the potential impact of surface-only LPRM data products is to mimic LSM transformations into integrated surface–root-zone products using a low-pass filter. Only marginal differences were detected between the weights obtained by using weekly surface and surface–root-zone mixed LPRM products (Fig. 2). Hence, overall these analyses suggest that differences in vertical support do not impact the analysis in a significant way.

6. Discussions and Conclusions

Model error covariance estimates in many hydrological data assimilation applications are obtained through perturbation of forcings and states without any rigorous justification of the magnitude of these perturbations [Reichle et al., 2008; Crow and van den Berg, 2010] even though the ensemble spread tends to be a stronger function of forcing spread than initial condition spread [Yilmaz et al., 2012]. Accordingly, this results in a merging scheme that is highly dependent on the user to accurately characterize modeling and observations errors which, in turn, determine the relative weight applied to model background and observations at update times.

In this study we have introduced a methodology that is completely objective and does not assume any arbitrary assumptions concerning the error characteristics of its input datasets. Specifically, error variances of three independently estimated soil moisture datasets were obtained using a triple collocation method and different soil moisture products were merged in an ordinary least squares framework. With the completely objective analysis introduced here, we are also able to estimate the uncertainty of the merged soil moisture as a separate product, which could be particularly useful for applications which require information about the reliability of the product.

The disadvantage of this framework when compared to traditional data assimilation techniques is that the estimated model errors are assumed to be stationary where in reality they could have time and/or flow dependency, and corrective information obtained via the merger is not temporally propagated forward in time (as in sequential filtering). Here it is stressed that we are not trying to replace the Kalman Filter based land data assimilation methodologies as they are more powerful than least squares merging through the ability of constraining all the model state and parameters with an adaptive error estimation framework. However, the least squares merging introduced here is more objective than many current land data assimilation applications in that it does not require any ad-hoc error estimates (i.e. forcing perturbations to create ensembles, observation error covariances).

There are three necessary assumptions in this methodology: the independence of errors, availability of long-enough time-series, and mutual linearity of products. The first assumption can be justified for many geophysical variables (i.e. soil moisture, soil temperature, potential evaporation, etc) as there are numerous independent satellite- and model-based estimates. However, currently there are no benchmarks or criteria established for the second assumption. Experience from synthetic simulations (results not shown) shows that the length of the available datasets used in this study may not be long enough to obtain highly accurate error estimates using triple collocation on weekly or monthly time-scales. On the other hand, Noah and LPRM [Owe et al., 2008] estimates for longer than two decades are already available (although they are not used in this study) and currently there are existing efforts to produce ALEXI estimates for similar time-periods. Additionally the availability of longer time-series will also enable estimating separate sets of weights for seasonal or sub-seasonal time-scales to partly address the issue of non-stationary weighting of products. The third assumption can be easily checked and the linearity can be justified via simple correlation calculations, as it is done in this study.

In this study we have applied a triple collocation-based merging strategy to integrate soil moisture information acquired from microwave remote sensing, thermal remote sensing and land surface modeling. The approach also provides the ability to estimate uncertainties associated with the merger estimate. When compared to ground-based soil moisture observations, our merged product improves upon the accuracy of its three parent products but fails to enhance merged products obtained using naive equal weighting. Given the small differences found between cross-correlations and weights, the lack of difference between our results and much naive weighting appears attributable to the marginal skill differences that exist between ALEXI, Noah, and LPRM based soil moisture estimates over the CONUS. We expect the differences between the skills of triple collocation- and naive method-based merged products would be higher over study areas where the differences between the skills of the parent products are higher.

Acknowledgments. We thank anonymous reviewers for their constructive comments, which led numerous clarifications in the final version of the manuscript. Research was partially supported by NASA Terrestrial Hydrology Program Grant NNX06AG07G entitled “Monitoring of Root-Zone Soil Moisture Using Multi-Frequency Observations of Surface Soil Moisture and Evapotranspiration”. USDA is an equal opportunity provider and employer.

References

Albergel, C., et al., 2008: From near-surface to root-zone soil moisture using an exponential filter: an assessment of the method based on in-situ observations and model simulations. *Hydrology and Earth System Sciences*, **12 (6)**, 1323–1337.

Anderson, M. C., C. Hain, B. Wardlow, A. Pimstein, J. R. Mecikalski, and W. P. Kustas, 2011: Evaluation of drought indices based on thermal remote sensing of evapotranspiration over the continental united states. *Journal of Climate*, **24 (8)**, 2025–2044.

Anderson, M. C., J. M. Norman, J. R. Mecikalski, J. A. Otkin, and W. P. Kustas, 2007a: A climatological study of evapotranspiration and moisture stress across the continental united states: 1. model formulation. *Journal of Geophysical Research*, **112**, D10 117.

Anderson, M. C., J. M. Norman, J. R. Mecikalski, J. A. Otkin, and W. P. Kustas, 2007b: A climatological study of evapotranspiration and moisture stress across the continental united states based on thermal remote sensing: 2. surface moisture climatology. *Journal of Geophysical Research*, **112**, D11 112.

Basara, J. B. and T. M. Crawford, 2000: Improved installation procedures for deep layer soil moisture measurements. *Journal of Atmospheric and Oceanic Technology*, **17**, 879–884.

Brock, F. V., K. C. Crawford, R. L. Elliott, G. W. Cuperus, S. J. Stadler, H. L. Johnson, and M. D. Eilts, 1995: The oklahoma mesonet: A technical overview. *Journal of Atmospheric and Oceanic Technology*, **12**, 5–19.

Caires, S. and A. Sterl, 2003: Validation of ocean wind and wave data using triple collocation. *Journal of Geophysical Research*, **108 (C3)**, 3098.

Cosh, M. H., T. J. Jackson, P. Starks, and G. Heathman, 2006: Temporal stability of surface soil moisture in the little washita river watershed and its applications in satellite soil moisture product validation. *Journal of Hydrology*, **323**, 168–177.

Crow, W. T., 2007: A novel method for quantifying value in spaceborne soil moisture retrievals. *Journal of Hydrometeorology*, **8**, 56–67.

Crow, W. T. and M. J. van den Berg, 2010: An improved approach for estimating observation and model error parameters in soil moisture data assimilation. *Water Resources Research*, **46**, W12 519.

Crow, W. T. and E. Van Loon, 2006: Impact of incorrect model error assumptions on the sequential assimilation of remotely sensed surface soil moisture. *Journal of Hydrometeorology*, **7**, 421–432.

Dai, Y., et al., 2003: The Common Land Model. *Bulletin of the American Meteorological Society*, **84 (8)**, 1013–1023.

Dorigo, W. A., K. Scipal, R. M. Parinussa, Y. Y. Liu, and W. Wagner, 2010: Error characterisation of global active and passive microwave soil moisture datasets. *Hydrology and Earth System Sciences*, **14**, 2605–2616.

Draper, C. S., J. P. Walker, P. J. Steinle, R. A. De Jeu, and T. R. Holmes, 2009: An evaluation of AMSR-E derived soil moisture over australia. *Remote Sensing of Environment*, **113 (4)**, 703 – 710.

Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayand, and J. D. Tarpley, 2003: Implementation of noah land surface model advances in the national centers for environmental prediction operational mesoscale eta model. *Journal of Geophysical Research*, **108 (D22)**, 8851.

Gauss, C., 1963: *Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections*. Dover, 326 pp.

Hain, C. R., W. T. Crow, J. R. Mecikalski, M. C. Anderson, and T. Holmes, 2011: An intercomparison of available soil moisture estimates from thermal infrared and passive microwave remote sensing and land surface modeling. *Journal of Geophysical Research*, **116**, D15 107.

Hain, C. R., J. R. Mecikalski, and M. C. Anderson, 2009: Retrieval of an available water-based soil moisture proxy from thermal infrared remote sensing. part i: methodology and validation. *Journal of Hydrometeorology*, **10**, 665–683.

Holmes, T. R. H., R. A. M. De Jeu, M. Owe, and A. J. Dolman, 2009: Land surface temperature from ka band (37 ghz) passive microwave observations. *Journal of Geophysical Research*, **114**, D04 113.

Illston, B. G., J. B. Basara, C. A. Fiebrich, K. C. Crawford, E. Hunt, D. K. Fisher, R. Elliott, and K. Humes, 2008: Mesoscale monitoring of soil moisture across a statewide network. *Journal of Atmospheric and Oceanic Technology*, **25 (2)**, 167–182.

Jackson, T. J., 1993: Measuring surface soil moisture using passive microwave remote sensing. *Hydrological Processes*, **7 (2)**, 139–152.

Kalman, R., 1960: A new approach to linear filtering and prediction problems. *Trans. ASME, J. Basic Eng.*, **82D**, 35–45.

Legendre, A., 1806: *Nouvelles Methodes pour la determination des orbites des cometes*. Courcier.

Loew, A. and F. Schlenz, 2011: A dynamic approach for evaluating coarse scale satellite soil moisture products. *Hydrology and Earth System Sciences*, **15**, 75–90.

Lu, H., T. Koike, H. Fujii, T. Ohta, and K. Tamagawa, 2009: Development of a physically-based soil moisture retrieval algorithm for spaceborne passive microwave radiometers and its application to AMSR-E. *Journal of the Remote Sensing Society of Japan*, **29 (1)**, 253–262.

Mecikalski, J. R., G. R. Diak, M. C. Anderson, and J. M. Norman, 1999: Estimating fluxes on continental scales using remotely sensed data in an atmospheric-land exchange model. *Journal of Applied Meteorology*, **38**, 1352–1369.

Miralles, D. G., W. T. Crow, and M. H. Cosh, 2010: Estimating spatial sampling errors in coarse-scale soil moisture estimates derived from point-scale observations. *Journal of Hydrometeorology*, **11**, 1423–1429.

Mladenova, I., V. Lakshmi, T. J. Jackson, J. P. Walker, O. Merlin, and R. A. De Jeu, 2011: Validation of AMSR-E soil moisture using l-band airborne radiometer data from national airborne field experiment 2006. *Remote Sensing of Environment*, **115 (8)**, 2096 – 2103.

Njoku, E. G. and S. K. Chan, 2006: Vegetation and surface roughness effects on AMSR-E land observations. *Remote Sensing of Environment*, **100**, 190–199.

Owe, M., R. De Jeu, and T. Holmes, 2008: Multisensor historical climatology of satellite-derived global land surface moisture. *Journal of Geophysical Research*, **113**, F01 002.

Owe, M., R. De Jeu, and J. P. Walker, 2001: A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index. *IEEE Transactions on Geoscience and Remote Sensing*, **39**, 1643–1654.

Parinussa, R., A. G. C. A. Meesters, Y. Y. Liu, W. Dorigo, W. Wagner, and R. de Jeu, 2011a: Error estimates for near-real-time satellite soil moisture as derived from the land parameter retrieval model. *Geoscience and Remote Sensing Letters, IEEE*, **8 (4)**, 779 –783.

Parinussa, R. M., T. R. H. Holmes, M. T. Yilmaz, and W. T. Crow, 2011b: The impact of land surface temperature on soil moisture anomaly detection from passive microwave observations. *Hydrology and Earth System Sciences*, **15**, 3135–3151.

Reichle, R. H., W. T. Crow, and C. L. Keppenne, 2008: An adaptive ensemble kalman filter for soil moisture data assimilation. *Water Resources Research*, **44 (3)**, W03 423.

Rodell, M., et al., 2004: The global land data assimilation system. *Bulletin of the American Meteorological Society*, **85 (3)**, 381–394.

Ryu, D. and J. S. Famiglietti, 2005: Characterization of footprint-scale surface soil moisture variability using gaussian and beta distribution functions during the southern great plains 1997 (sgp97) hydrology experiment. *Water Resources Research*, **41**, W12 433.

Schaefer, G. L., M. H. Cosh, and T. J. Jackson, 2007: The usda natural resources conservation service soil climate analysis network (scan). *Journal of Atmospheric and Oceanic Technology*, **24**, 0739–0572.

Scipal, K., T. Holmes, R. de Jeu, V. Naeimi, and W. Wagner, 2008: A possible solution for the problem of estimating the error structure of global soil moisture data sets. *Geophysical Research Letters*, **35**, L24 403.

Sorenson, H. W., 1970: Least-squares estimation: from gauss to kalman. *IEEE spectrum*, **7**, 63–68.

Stoffelen, A., 1998: Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. *Journal of Geophysical Research*, **103 (C4)**, 7755–7766.

Talagrand, O., 1997: Assimilation of observations: An introduction. *Journal of the Meteorological Society of Japan*, **75**, 191–209.

Wagner, W., G. Lemoine, and H. Rott, 1999: A method for estimating soil moisture from ers scatterometer and soil data. *Remote Sensing of Environment*, **70 (2)**, 191 – 207.

Yilmaz, M. T., T. Delsole, and P. R. Houser, 2012: Reducing water imbalance in land data assimilation: Ensemble filtering without perturbed observations. *Journal of Hydrometeorology*, **13 (1)**, 413–420, doi:10.1175/JHM-D-11-010.1.

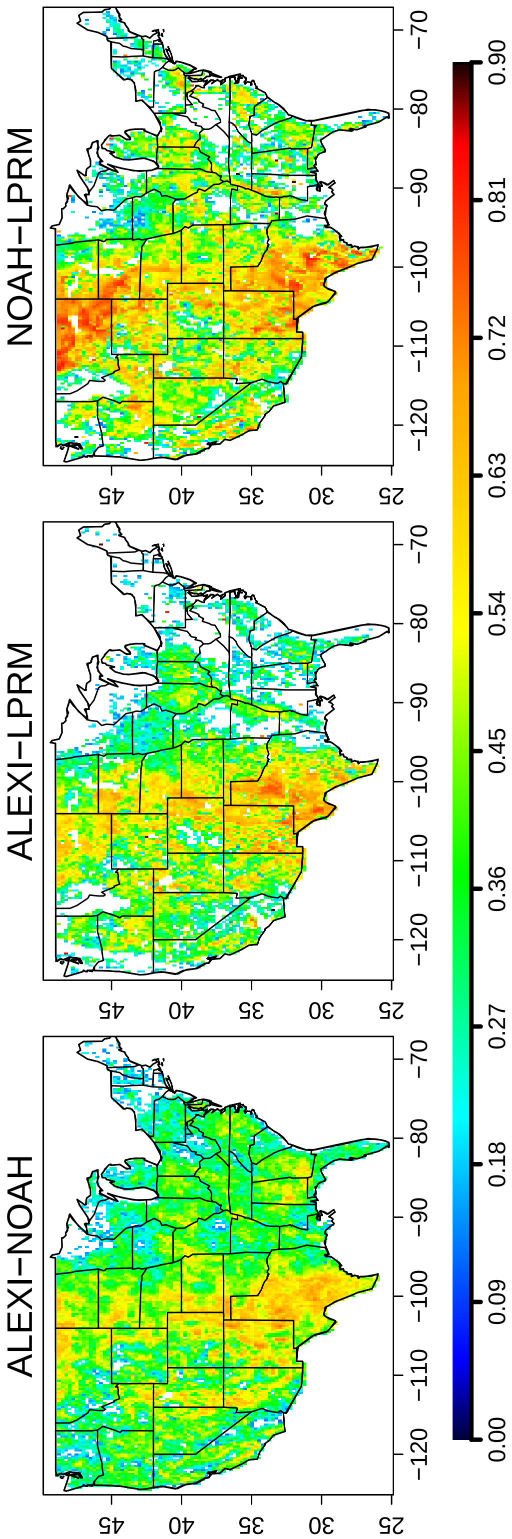

Figure 1. Cross-correlations (r^2) between weekly ALEXI, Noah, and LPRM composites during 2002-2010 using months April through October.

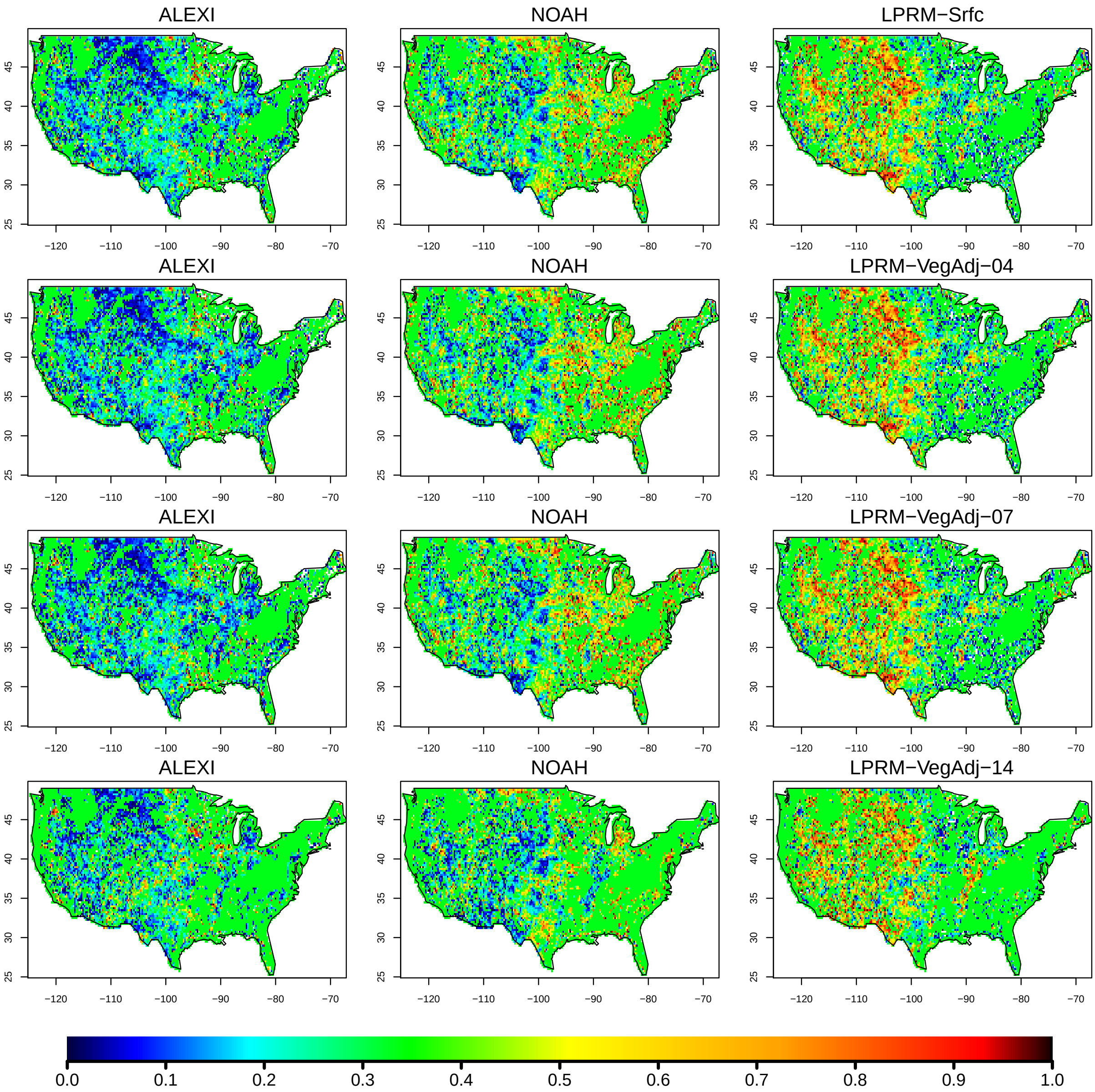

Figure 2. Weights of soil moisture estimates obtained from triple collocation. All four rows used the same ALEXI and Noah products in the triple collocation analysis. The first row used the native LPRM surface soil moisture product, whereas the second to fourth rows used also the exponentially filtered LPRM-based root-zone soil moisture products with characteristic time-lengths of 4, 7, and 14 days respectively. Here the areas over where triple collocation analyses were not applied due to data unavailability were assigned 0.33 weight for all three products.

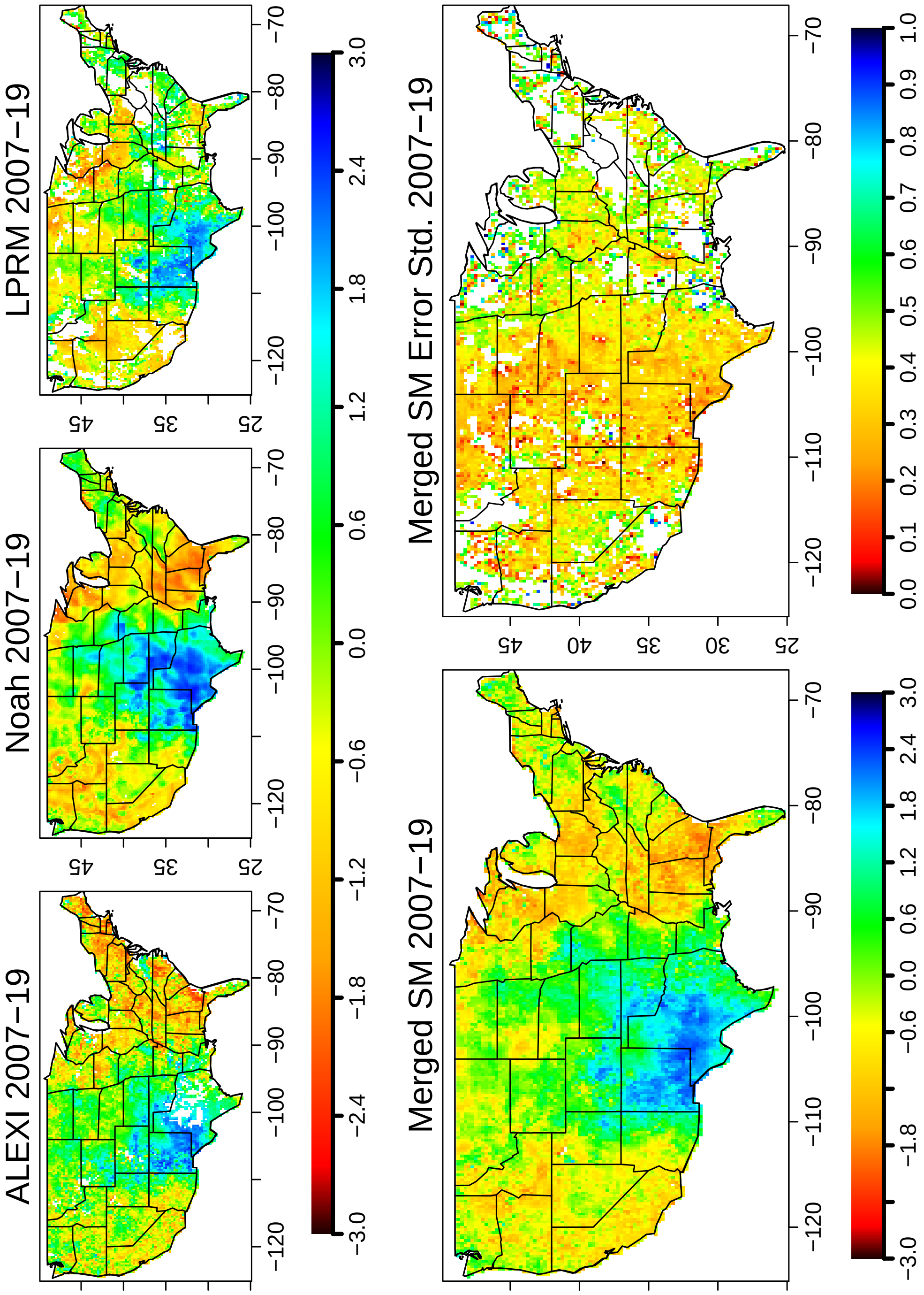

Figure 3. Weekly composites of ALEXI, Noah, LPRM, merged soil moisture and its uncertainty estimates for the 19th week of 2007. Soil moisture estimates are presented in terms of standard normal deviates.

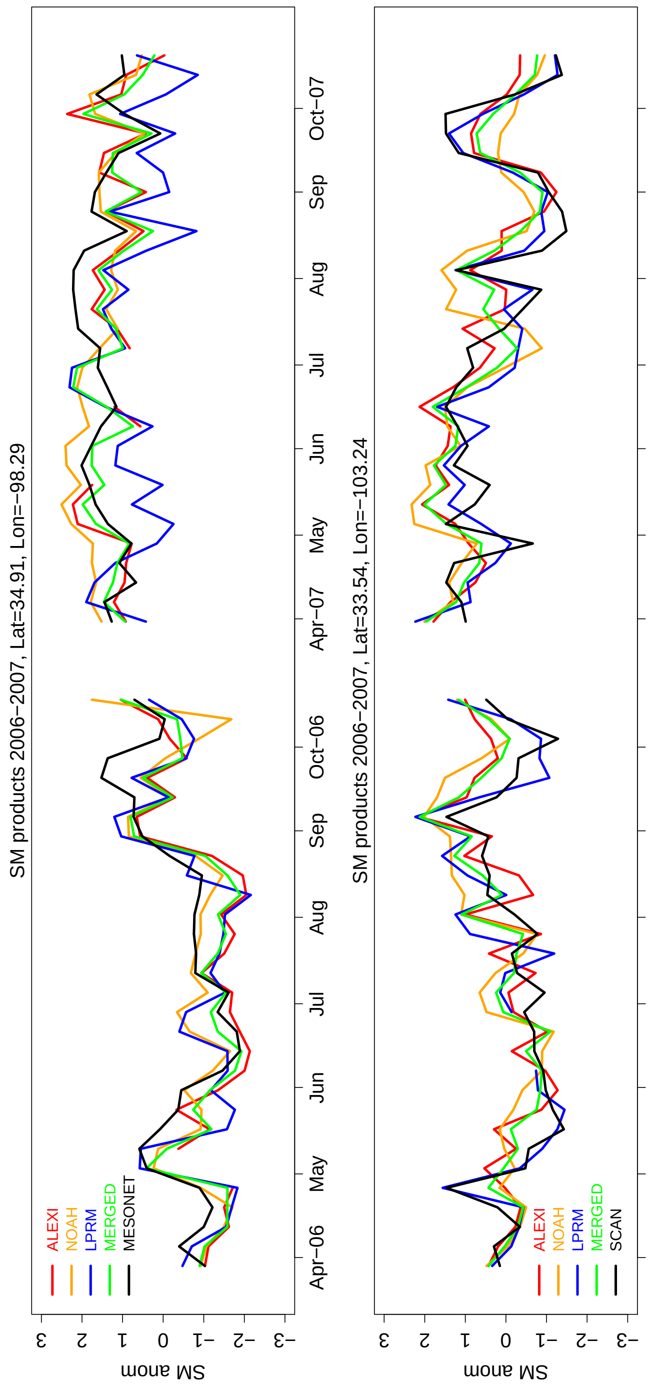

Figure 4. Weekly soil moisture composite time series in terms of standard normal deviates. Upper and lower panels correspond to time series at one of MESONET (Apache) and SCAN (Crossroads) stations respectively. ALEXI, Noah, and LPRM values are obtained from the closest available station.

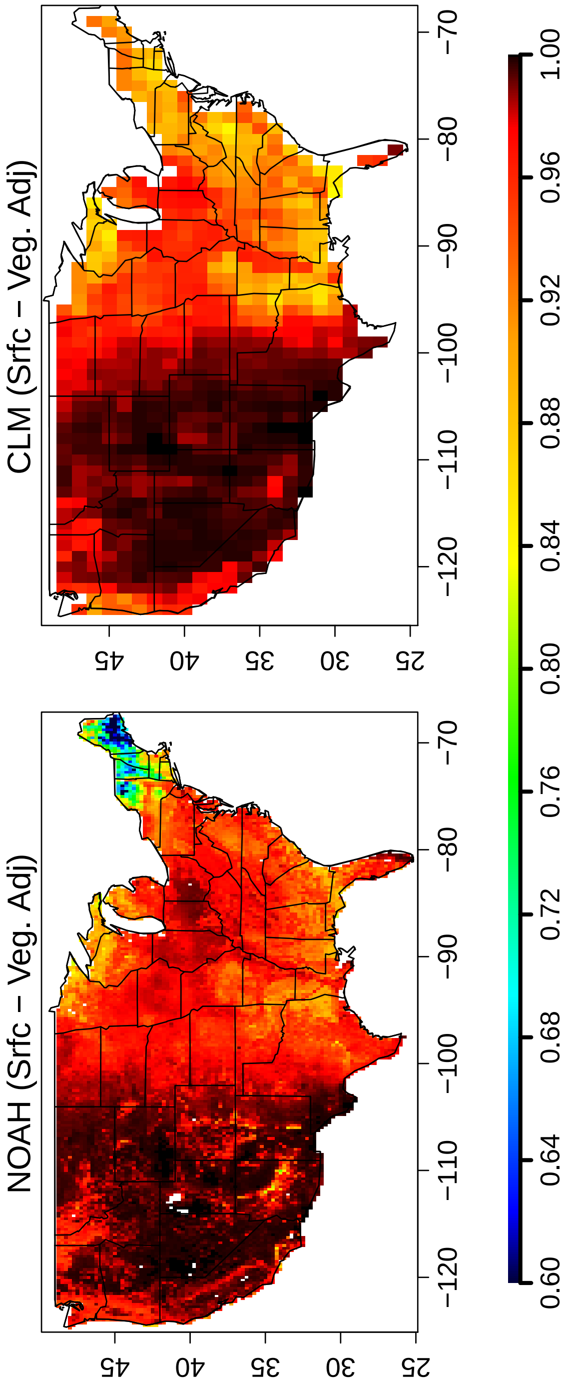

Figure 5. Weekly composite correlations between surface and vegetation-adjusted soil moisture estimates of Noah and CLM over the CONUS.

Table 1. Parent products (ALEXI, Noah, LPRM), merged estimate, and station data (MESONET or SCAN) cross-correlations with the station data. Three layers of station soil moisture data are considered: surface, vegetation-adjusted, and root-zone. NAIVE refers to the merged product obtained by giving equal weight to each parent products.

	MESONET			SCAN		
	Surface	Veg. Adj.	Root	Surface	Veg. Cor.	Root
ALEXI	0.46	0.48	0.38	0.36	0.38	0.34
Noah	0.54	0.54	0.33	0.41	0.42	0.33
LPRM	0.52	0.55	0.43	0.51	0.54	0.51
MERGED	0.61	0.63	0.46	0.55	0.58	0.51
NAIVE	0.61	0.64	0.46	0.55	0.57	0.50
MESONET or SCAN (Surface)	1.00	0.91	0.37	1.00	0.91	0.67
MESONET or SCAN (Veg. adj.)	0.91	1.00	0.60	0.91	1.00	0.78
MESONET or SCAN (Root)	0.37	0.60	1.00	0.67	0.78	1.00

Table 2.

Results of product versus ground-data cross-correlation analysis for various scenarios. "Total" refers to the number of ground-stations considered. Neg and Pos refer to statistically-significant negative and positive results respectively for the scenarios given in the left column, and Non refers to neither a positive result or a negative result. For the significance tests, a 95% confidence level is used.

		MESONET				SCAN			
Scenario	Product	Total	Neg	Non	Pos	Total	Neg	Non	Pos
Correlations	ALEXI	51	0	2	49	50	0	5	45
significantly	Noah	51	0	1	50	50	0	4	46
different	LPRM	50	0	1	49	44	0	7	37
than 0	MERGED	51	0	0	51	50	0	2	48
Merged estimate	ALEXI	51	5	-	46	50	4	-	46
correlations better	Noah	51	12	-	39	50	19	-	31
than individual products	LPRM	50	4	-	46	44	7	-	37
(no significance test)									
Naive estimate	ALEXI	51	3	-	48	50	3	-	47
correlations better	Noah	51	10	-	41	50	23	-	27
than individual products	LPRM	50	7	-	43	44	10	-	34
(no significance test)									
Merged best significantly	ALL	51	0	50	1	50	0	50	0
Merged best	ALL	51	0	19	32	50	0	29	21
Naive best significantly	ALL	51	0	48	3	50	0	49	1
Naive best	ALL	51	0	18	33	50	0	33	17

Table 3. Mean weights and cross-correlations over the CONUS for different data compositing strategies.

	Weights		
	ALEXI	Noah	LPRM
Mutually available weekly	0.27	0.35	0.41
Mutually available monthly	0.34	0.32	0.37
All available weekly	0.25	0.41	0.37
All available monthly	0.32	0.37	0.35
	Correlations		
	ALEXI-Noah	ALEXI-LPRM	Noah-LPRM
Mutually available weekly	0.38	0.40	0.43
Mutually available monthly	0.44	0.45	0.45
All Available weekly	0.40	0.38	0.44
All Available monthly	0.46	0.44	0.46

Table 4. Noah, CLM, and station cross-correlations between surface and vegetation-adjusted weekly soil moisture composite values at multiple locations. CONUS-East lays between 88°-75°W, and 32°-41°N and CONUS-West lays between 116°-103°W and 29°-36°N.

Surface – Veg. Adj.	MESONET Stations	SCAN Stations	CONUS	CONUS-East	CONUS-West
Noah	0.95	0.96	0.96	0.96	0.99
CLM	0.96	0.96	0.96	0.92	0.99
MESONET	0.91	-	-	-	-
SCAN	-	0.91	-	-	-